Overview of Evaluation

BACKGROUND

This section includes activities that address

- Defining evaluation
- Using evaluation logic
- Understanding different types of evaluation
- Exploring the history of evaluation
- Understanding the differences between evaluation and research

The following information is provided as a brief introduction to the topics covered in these activities.

DEFINITION OF EVALUATION

Various definitions of evaluation have been offered over the years. While each definition takes a slightly different view of evaluation, they all share important commonalities. First, evaluation is viewed as a systematic process. It should not be conducted as an afterthought; rather, it is a planned and purposeful activity. Second, evaluation involves collecting data regarding questions or issues about society in general and organizations and programs in particular. Third, evaluation is a process for enhancing knowledge and decision making, whether the decisions are for improving or refining a program, process, product, system, or organization or for determining whether

or not to continue or expand a program. And, in each of these decisions, there is some aspect of judgment about the merit, worth, or value of the evaluand (that which is being evaluated). Finally, the notion of evaluation use is either implicit or explicit in most definitions. Ultimately, evaluation is concerned with asking questions about issues that arise out of everyday practice. It is a means for gaining better understanding of what we do and the effects of our actions in the context of society and the work environment. A distinguishing characteristic of evaluation is that, unlike traditional forms of academic research, evaluation is grounded in the everyday realities of organizations. Evaluations can be conducted of programs, processes, products, systems, organizations, personnel, and policies.

**EVALUATION LOGIC**

One distinguishing feature of evaluation is that it has a particular logic that supports evaluation practice. This logic refers to specific reasoning principles to support decisions regarding how one determines the processes and impact of a program (or that which is being evaluated). The logic of evaluation generally involves the following actions:

- *Establishing Criteria.* On what dimensions must it do well?
- *Constructing Standards.* How well should it perform on each dimension?
- *Measuring Performance and Comparing With Standards.* How well did it perform on each dimension?
- *Synthesizing and Integrating Evidence Into a Judgment of Worth.* What is its merit, worth, or value?
- *Making Recommendations.* What recommendations should be made?

Following the logic of evaluation helps evaluators and program staff understand the steps one undertakes to come to an evaluative judgment about the merit, worth, or value of the evaluand.

**TYPES OF EVALUATION**

Evaluation is often conducted to gain information before a program’s development, as in a needs analysis; to improve and refine a program; or to make judgments about a program’s future. The following are descriptions of the different types of evaluation.

- *Developmental evaluation* positions the evaluator as a part of a program’s design and development process. The evaluator collects
information and provides informal feedback to members of the
design team and possibly organization members to help them
perfect the program being designed before it is ready for beta or pilot
testing.
• *Formative evaluation* is typically conducted for the purpose of
improvement.
• *Summative evaluation* is implemented for the purpose of determining
the merit, worth, or value of the evaluand in a way that leads to
making a final evaluative judgment. It is usually conducted after a
program’s completion.

## HISTORY OF EVALUATION

The development of the evaluation profession is described in the following
highlights.

• Until the late 1950s and early 1960s, evaluation mainly focused on
educational assessment and was conducted by social science researchers in
a small number of universities and organizations.

• President Lyndon Johnson’s War on Poverty and the Great Society
programs of the 1960s spurred a large investment of resources in social and
educational programs. Senator Robert Kennedy, concerned that the federal
money would be misspent and not used to help disadvantaged children,
delayed passage of the Elementary and Secondary Education Act (ESEA)
until an evaluation clause was included. The resulting bill required the local
education agency to submit an evaluation plan, and the state agency, a sum-
mary report. Consequently, evaluation requirements became part of every
federal grant. Early expectations of evaluation were that it would illuminate
the causes of social problems and the clear and specific means with which to
fix such problems.

• Two U.S.-based professional evaluation associations emerged in 1976:
the Evaluation Network, which mostly consisted of university professors and
school-based evaluators, and the Evaluation Research Society, which
attracted primarily government-based and university evaluators. In 1985,
these two organizations merged to form the American Evaluation Associa-
tion (AEA; www.eval.org), whose membership includes approximately
3,700 individuals from across the globe.

• Throughout the 1970s and 1980s, growing concerns were voiced
about the utility of evaluation findings and the use of experimental and
quasi-experimental designs.
• In the 1980s, huge cuts in social programs resulted from President Ronald Reagan's emphasis on less government involvement. The requirement to evaluate was removed or lessened from many federal grants.

• During the 1980s, many school districts, universities, private companies, state departments of education, the Federal Bureau of Investigation (FBI), the Food and Drug Administration, and the General Accounting Office (GAO) developed internal evaluation units.

• In the 1990s, there was an increased emphasis on government program accountability and organizations' efforts to be lean, efficient, global, and more competitive. Evaluation was conducted not only to meet government mandates but also to improve programs' effectiveness, to enhance organizational learning, and to make resource allocation decisions in a wide variety of both public and private organizations. In addition, an increasing number of foundations created internal evaluation units, provided support for evaluation activities, or both.

• The reauthorization of ESEA that resulted in the No Child Left Behind (NCLB) Act of 2001 (signed into law on January 8, 2002, by President George W. Bush) is considered the most sweeping reform of elementary and secondary education since 1965. It has redefined the federal role in K–12 education by focusing on closing the achievement gap between disadvantaged and minority students and their peers. Its four basic principles include (1) stronger accountability for results, (2) increased flexibility and local control, (3) expanded options for parents, and (4) an emphasis on proven teaching methods. This legislation has had a profound influence on educational evaluation designs and methods by emphasizing the use of randomized control group experiments. Concern over this situation led to the development of a response to the U.S. Department of Education by the AEA in 2003. The AEA's response expressed concern that the legislation "manifests fundamental misunderstandings about (1) the types of studies capable of determining causality, (2) the methods capable of achieving scientific rigor, and (3) the types of studies that support policy and program decisions" (American Evaluation Association, 2003).

• At present, there is increasing interest in participatory, collaborative, and learning-oriented formative evaluations in lieu of summative evaluations conducted by the lone evaluator. Evaluators readily accept that few evaluations are value-free, and, indeed, most are politically charged.

• National evaluation associations are becoming established throughout the world. These include, but are not limited to, the Australasian Evaluation Society, the African Evaluation Association, the Canadian Evaluation Society, the European Evaluation Society, the United Kingdom Evaluation Society, the Malaysian Evaluation Society, the Central American Evaluation Association, and the Japan Evaluation Society.
EVALUATION VERSUS RESEARCH

The terms evaluation and research are sometimes used interchangeably. Although the two forms of inquiry use the same data collection and analysis methods, they differ significantly in several ways.

- They have different purposes.
- They respond to different audiences’ information needs.
- They pose different kinds of questions.
- They communicate and report their findings in different ways and to different groups.
- They have different expectations regarding the use of the results.

Although it is important to understand the difference between these two forms of inquiry, some may view this difference as merely technical and not particularly practical. What is important is the inquiry itself—the collection of quality information that will inform and guide decision making, learning, and action.

RESOURCES


Activity 1
Images and Purposes of Evaluation

Overview
This activity helps participants articulate their current perceptions of evaluation: what it is, why it is done, and how it is performed. It also provides an overview of the different types of evaluation.

Instructional Objectives
Participants will
- Reflect on their assumptions and experiences regarding evaluation
- Learn how different people perceive the value of evaluation
- Understand the various definitions, purposes, and types of evaluation

Number of Participants
- Minimum number of participants: 3
- Maximum number of participants: unlimited

Time Estimate: 75 to 120 minutes
There are three components to this activity. The first one, which asks participants to draw a picture of evaluation, requires approximately 30 to 45 minutes. The second component, which involves a discussion of evaluation definitions, requires approximately 15 to 30 minutes. The third component, which involves a discussion of purposes and types of evaluation, can be accomplished in 30 to 45 minutes.

Materials Needed
- Pens/pencils
- Paper for drawing
- Flipchart, markers, tape
• Handout “Definitions of Evaluation”
• Handout “Types of Evaluation”

**Instructional Method**

Visual representation

**Procedures**

*Part 1 (30 to 45 minutes)*

Facilitator’s tasks:

• Distribute blank pieces of paper and markers or other writing utensils.
• Ask participants to draw the first image that comes to mind when they hear or see the word *evaluation*. Ask participants to label their drawings with one or two words that describe or represent the image. Tell participants they have 3 to 4 minutes to do this.
• Ask participants to get into groups of three to five people and share their images, identify common themes in their images, and write their group’s themes on a piece of flipchart paper.
• Invite participants to share their group’s themes.
• Refer to the themes and images while leading a discussion about individuals’ common perceptions of evaluation and the implications these perceptions may have on evaluation practice. Ask participants the following questions:
  – Why do you think you drew that particular image? What experiences have you had that influenced what you drew?
  – When individuals perceive evaluation as a negative endeavor, what consequences might this have for conducting evaluation within organizations? Conversely, what implications do positive images have for conducting evaluations within organizations?

*Part 2 (15 to 30 minutes)*

Facilitator’s tasks:

• Distribute the handout “Definitions of Evaluation” and ask the following questions to the entire group:
  – What are some commonalities among these definitions?
  – How do these definitions differ?
  – Which of these definitions resonate with your own evaluation experiences?
  – Are any of these definitions reflected in your images of evaluation?
  – How important do you think it is to define evaluation?
Part 3 (30 to 45 minutes)

Facilitator’s tasks:

- Distribute the handout “Types of Evaluation.”
- Ask participants to get into pairs to complete the worksheet.
- Invite participants to share some of the questions they developed for each type of evaluation. Write their questions on the flipchart.
- Debrief the activity with the following questions:
  - What new insights did you gain from these activities?
  - How would you now describe evaluation to someone?
  - How is your current understanding of evaluation different from your understanding before this activity?
Definitions of Evaluation

Handout for Activity 1

Several definitions of evaluation have been offered over the years. Michael Scriven (1967, 1991) developed one of the earliest definitions that is still commonly used today:

Evaluation refers to the process of determining the merit, worth, or value of something, or the product of that process. Terms used to refer to this process or part of it include: appraise, analyze, assess, critique, examine, grade, inspect, judge, rate, rank review, study, test. . . . The evaluation process normally involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of evaluands on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations. (Scriven, 1991, p. 139)

Another definition, which stems from evaluation’s long history with social programs and takes on a social science research perspective, comes from Rossi, Lipsey, and Freeman (2004):

Program evaluation is the use of social research methods to systematically investigate the effectiveness of social intervention programs. It draws on the techniques and concepts of social science disciplines and is intended to be useful for improving programs and informing social action aimed at ameliorating social problems. (p. 28)

A definition used by many evaluation practitioners is from Patton (1997), who emphasizes the use of evaluation findings:

Program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming. (p. 23)

Preskill and Torres (1999) offer a definition that focuses on evaluative activities specifically conducted within organizations for the purpose of organizational learning and change:

We envision evaluative inquiry as an ongoing process for investigating and understanding critical organization issues. It is an approach to learning that is fully integrated with an organization’s work practices, and as such, it engenders (a) organization members’ interest and ability in exploring critical issues using evaluation logic, (b) organization members’ involvement in evaluative processes, and (c) the personal and professional growth of individuals within the organization. (pp. 1–2)
References


## Types of Evaluation

### Handout for Activity 1

<table>
<thead>
<tr>
<th>Types of Evaluation</th>
<th>Questions an Evaluation Would Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developmental evaluation</td>
<td>positions the evaluator as a part of a program’s design and development process.</td>
</tr>
<tr>
<td>Formative evaluation</td>
<td>is typically conducted for the purpose of refining or improving a program and is often conducted by internal evaluators.</td>
</tr>
<tr>
<td>Summative evaluation</td>
<td>is implemented for the purpose of determining the merit, worth, or value of the evaluand in a way that leads to a final evaluative judgment. It is usually conducted after completion of the program (for ongoing programs) and for the benefit of some external audience or decision makers. External evaluators often conduct summative evaluations.</td>
</tr>
</tbody>
</table>
**Monitoring and auditing** are often associated with the need to determine if a program is being administered in ethical or legal ways; it is a means for checking the program’s implementation. A monitoring type of evaluation would focus on the extent to which program administrators are wasting funds, inappropriately using staff resources, or ineffectively tracking participants’ involvement in the program.

**Outcome evaluation** seeks to understand intended changes in knowledge, attitudes, and practices that result from a program or project’s intervention.

**Impact evaluation** focuses on what happens to participants because of the intervention or program.

**Performance management** is typically undertaken within governments and the nonprofit sector to meet a demand for documentation of results. It focuses on program activities (process), direct products and services delivered by a program (outputs), and the results of those product and services (outcomes).
Activity 2

Applications of Developmental, Formative, and Summative Evaluations

Overview

This activity helps participants understand the purposes and uses of developmental, formative, and summative approaches to evaluation.

Instructional Objectives

Participants will

- Consider when and how to frame an evaluation as developmental, formative, or summative
- Identify the purposes and associated questions for an evaluation that may have a developmental, formative, or summative focus
- Understand the implications and issues related to conducting developmental, formative, and summative evaluations

Number of Participants

- Minimum number of participants: 3
- Maximum number of participants: unlimited when participants are in groups of 3 to 5

Time Estimate: 30 to 45 minutes

In addition to providing the necessary background information on developmental, formative, and summative approaches to evaluation, this activity requires approximately 30 to 45 minutes, depending on the number of participants (or groups) and the time available for discussion.
Materials Needed

- Pens/pencils
- Flipchart, markers, tape
- Handout “Developmental, Formative, and Summative Evaluations: Purposes and Questions”

Instructional Method

Case scenario

Procedures

Facilitator’s tasks:

- Ask participants to get into groups of three to five people.
- Distribute the handout “Developmental, Formative, and Summative Evaluations: Purposes and Questions.”
- Instruct participants to discuss and complete the handout and to transfer their work to a piece of flipchart paper.
- Ask participants to attach their flipchart page to a wall.
- Invite participants to discuss the purposes and questions they developed.
- Debrief the activity with the following questions:
  - What issues did this activity raise for you?
  - How would an organization choose whether to do a developmental, formative, or summative evaluation (or all of these)?
  - What are the implications of implementing each of these approaches to evaluation?
  - Why might an organization conduct only a summative evaluation?
  - Some might say that conducting a developmental evaluation is not an appropriate role for an evaluator. What is your response to this?
A school district wants to provide leadership training and development opportunities to its administrators. They ask the Professional Staff Development Department (PSDD) to develop a leadership development program for the district’s 90 principals, assistant principals, and department heads.

### Developmental Evaluation

As the PSDD designs the program, what kinds of evaluative feedback should it seek?

### Formative Evaluation

The program has been designed, and the first 30 administrators have now completed the program. What kinds of evaluative feedback should the PSDD seek?

### Summative Evaluation

The program has been offered to all of the district’s 90 administrators over the last three years. What kinds of evaluative feedback should the PSDD seek?
Activity 3

Evaluating Chocolate Chip Cookies Using Evaluation Logic

Overview

In this activity, participants take a normal, everyday experience, such as evaluating chocolate chip cookies, and learn what it means to make evaluative criteria, standards, and judgments explicit.

Instructional Objectives

Participants will

• Understand the concept of evaluation logic and how it is the foundation for evaluation practice
• Use evaluation logic to evaluate chocolate chip cookies
• Describe how evaluation logic applies to evaluating programs, services, policies, procedures, systems, and organizations

Number of Participants

• Minimum number of participants: 3
• Maximum number of participants: unlimited when participants are in groups of 3 to 5

Time Estimate: 45 to 90 minutes

This activity requires approximately 45 to 90 minutes, depending on the number of participants (or groups) and the time available for discussion.

Materials Needed

• Pens/pencils
• Four different kinds of chocolate chip cookies. The cookies should be different brands yet of the same type. For example, all of the cookies should have either milk chocolate chips or semisweet chips, and all of them should be without nuts or coconut. If you wish to introduce the
notion of confounding variables, then you might include a cookie that has nuts, coconut, or another ingredient not represented in the other cookies.

- Plastic bags in which to put the cookies
- Plastic or paper plates
- Napkins
- Handout “Evaluating Chocolate Chip Cookies Using Evaluation Logic”

**Instructional Method**

Small-group work

**Procedures**

*Prior to Class*

Facilitator’s tasks:

- Purchase four different brands of chocolate chip cookies and paper napkins and plates.
- Determine how many groups will participate and obtain four small plastic bags for each group.
- Break cookies into bite-size pieces and place each brand into separate plastic storage bags, one bag for each brand. Label the bags \(A\), \(B\), \(C\), and \(D\). Repeat this for each group. For example, if four groups are participating, you need four bags of cookie \(A\), four bags of cookie \(B\), four bags of cookie \(C\), and four bags of cookie \(D\).

*During Class*

Facilitator’s tasks:

- Tell participants that they will be conducting an evaluation of chocolate chip cookies. Explain that this is an exercise in understanding the underlying logic of evaluation.
- Ask participants to get into groups of three to five people.
- Provide each group with four cookie bags (labeled \(A\), \(B\), \(C\), and \(D\); each group now has four bags with a different brand of cookie in each bag). Give each group napkins and paper plates on which to put their cookies when they are ready for the taste test.
- Ask participants to keep notes on their process, the decisions they made, and any challenges they experienced.
- Distribute the handout “Evaluating Chocolate Chip Cookies Using Evaluation Logic.”
• Tell participants that if they do not wish to eat the cookies (due to health, diet, or other reasons), they do not have to do so. Ask them to be participant observers.
• Explain that they are to first discuss and complete the first two columns on the handout—the criteria for which chocolate chip cookies should be judged and the standards that should be used to judge a chocolate chip cookie.
• Tell participants to then taste each cookie and complete the third and fourth columns on the handout.
• Debrief the activity with the following questions:
  – How would describe your experience in establishing criteria for evaluating the cookies? Was it difficult? What were some of the things you discussed in your group?
  – How did you determine what standards to use? Were you all in agreement on these standards? How did you reconcile your differences?
  – How comfortable were you with your final judgment about which cookie was best and which you would recommend?
  – For those of you who did not taste the cookies, what role did you play in this evaluative process? What did you observe about your group’s behavior? What impact did you have on your group’s evaluative judgment and recommendation?
# Evaluating Chocolate Chip Cookies Using Evaluation Logic

## Handout for Activity 3

<table>
<thead>
<tr>
<th>Establishing Criteria</th>
<th>Constructing Standards</th>
<th>Measuring Performance and Comparing With Standards</th>
<th>Synthesizing and Integrating Evidence Into Judgment of Worth</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the dimensions or criteria on which a chocolate chip cookie should be judged (e.g., type of chocolate, texture, color, aroma)?</td>
<td>How well should the cookie perform on each of the dimensions (what are your standards)? Develop a rating system.</td>
<td>Based on your criteria for a good cookie, how well does each cookie measure up against the standards you set?</td>
<td>Which cookie is worth buying?</td>
</tr>
</tbody>
</table>

Which cookie would you recommend?
Activity 4
Tracing Evaluation’s History

Overview
This activity helps participants understand the origins of modern-day evaluation practice and the various contexts in which it has evolved into a thriving profession.

Instructional Objectives
Participants will

- Discuss the origins of the evaluation profession beginning in the 1960s
- Develop a timeline that highlights key points in the profession’s development
- Identify and discuss the various cultural, societal, economic, and political influences on evaluation practice since the 1960s

Number of Participants

- Minimum number of participants: 3
- Maximum number of participants: unlimited when participants are in groups of 3 to 5

Time Estimate: 45 to 60 minutes
In addition to providing the necessary background on the history of evaluation, this activity requires approximately 45 to 60 minutes, depending on the number of participants (or groups) and the time available for discussion.

Materials Needed

- Pens/pencils
- Flipchart, markers, tape
- Handout “Tracing Evaluation’s History”
Instructional Method
Visual representation

Procedures
Facilitator’s tasks:

• Ask participants to get into groups of three to five people.
• Distribute the handout “Tracing Evaluation’s History.”
• Instruct participants to develop a timeline that represents the development of the evaluation profession by including key cultural, societal, economic, and political milestones.
• Ask each group to attach their completed timeline to a wall.
• Invite participants to describe their timelines.
• Debrief the activity with the following questions:
  – What insights have you gained from this activity regarding the evaluation profession’s development? What surprised you? What do you know now that you did not know before?
  – What do you think has had the greatest influence on the profession’s development?
  – Based on where evaluation has been, where do you think the profession is heading? What might be new areas of development?
  – What do you think evaluation will look like in 10 years? What might influence the practice of evaluation in the next few years?
Tracing Evaluation’s History

Handout for Activity 4

Directions:

1. Discuss what you believe are key dates in the evaluation profession’s development.

2. Identify the cultural, societal, economic, and political influences that might have influenced the profession’s development at these points in time.

3. Develop a timeline that depicts these dates and influences.
Activity 5
Evaluation Versus Research

Overview
This activity helps participants understand the similarities and differences between evaluation and research.

Instructional Objectives
Participants will

- Identify the similarities and differences between evaluation and research as forms of inquiry
- Understand why evaluation is often confused with research
- Be able to articulate the differences between evaluation and research

Number of Participants
- Minimum number of participants: 4
- Maximum number of participants: unlimited when participants are in groups of 3 to 5

Time Estimate: 45 to 60 minutes
In addition to providing the necessary background information on the differences between evaluation and research, this activity requires approximately 45 to 60 minutes, depending on the number of participants (or groups) and the time available for discussion.

Materials Needed
- Pens/pencils
- Flipchart, markers, tape
- Handout “Evaluating and Researching the Learning Pill”
**Instructional Method**

Case scenario

**Procedures**

*Prior to Class*

Facilitator’s tasks:

- Purchase a large bag of M&M candies.
- Purchase enough small plastic bags or cups so that each participant has one.
- Place 10 M&Ms into each bag or cup.

*During Class*

Facilitator’s tasks:

- Ask participants to get into groups of three to five people.
- Announce that one or more of the groups are “evaluators” and the other groups are “researchers.”
- Hand out one bag or cup of M&Ms to each participant.
- Distribute the handout “Evaluating and Researching the Learning Pill.”
- Provide the following directions:

  You have just read of a new discovery that could potentially revolutionize learning and development. A group of pharmaceutical researchers has developed a learning pill. Coincidentally, soon after reading about this discovery, you receive a call from one of the pill’s developers who asks you to help them study the effects of this learning pill. The person explains that her company would like you and your colleagues to conduct an evaluation of the pill and conduct research on the pill. Because of the high development costs associated with this learning pill, you each have been given only 10 pills to use in your studies, which you may use in any way you wish. Your task is to design each of these inquiries using the handout.

- Instruct participants to complete the handout.
- After the specified time for the small-group work, ask the evaluator group(s) to meet with the researcher group(s) to do the following:
  - Compare the results from their handouts.
  - Discuss similarities in their designs.
  - Discuss differences in their designs.
• Debrief the activity with the following questions:
  – What did you learn about the differences and similarities of these two forms of inquiry?
  – What surprised you? What hadn’t you thought of before?
  – To what extent do you think that any distinction between evaluation and research is artificial?
  – How would you now describe the differences between research and evaluation to someone who used the terms interchangeably?
Evaluating and Researching the Learning Pill

Handout for Activity 5

Directions:
Circle whether your group represents evaluators or researchers in the first column. Write your answers to each of the questions in the second column.

<table>
<thead>
<tr>
<th>Evaluation or Research Design Component</th>
<th>Evaluation or Research Design Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the purpose of the study?</td>
<td></td>
</tr>
<tr>
<td>Who are the intended users of the study's findings?</td>
<td></td>
</tr>
<tr>
<td>What is the focus of the study?</td>
<td></td>
</tr>
<tr>
<td>What methods would you use to collect data (e.g., tests, interviews, surveys, observation)?</td>
<td></td>
</tr>
<tr>
<td>How would you report the results of your study, and to whom?</td>
<td></td>
</tr>
<tr>
<td>If your study found that the pill had no effect, would the findings still be reported? Explain your reasons.</td>
<td></td>
</tr>
</tbody>
</table>
